Economics and Sociology
Occasional Paper No. 1870

CONTEMPORARY PRODUCTION THEORY: ITS IMPORTANCE IN AGRICULTURAL RESEARCH

by

Nelson Aguilera-Alfred

December 1990

Agricultural Finance Program
Department of Agricultural Economics
and
Rural Sociology
The Ohio State University
2120 Fyffe Road
Columbus, Ohio 43210-1099

Abstract

This paper discusses in details several important aspects of contemporary production theory and its relationship to neoclassical theory. In particular, it provides a detailed presentation of fundamentals of duality, multiproduct cost functions, elasticity of substitution between input pairs in its various forms, production factors functional separability, functional forms, and the concept of factor price frontier. The approaches outlined in this paper have applications to studies conducted for entire regions or countries, but is also applicable to studies conducted on data from farm records for individual firms.

CONTEMPORARY PRODUCTION THEORY: ITS IMPORTANCE IN AGRICULTURAL RESEARCH

I. Fundamentals of Duality

Agricultural economists are perhaps most familiar with the concept of duality as it relates to linear programming models. In these models, duality refers to the fact that any linear programming model can be expressed either as a maximization (minimization) problem, the **primal**, or a corresponding minimization (maximization) problem, the **dual**, subject to appropriate linear constraints. The key characteristic of this dual relationship is that all the information about the solution to the primal may be obtained from the corresponding dual. Thus, by solving the dual problem all the information regarding its solution may be obtained without resolving the primal problem itself. Contemporary production theory focuses on the conditions under which a dual relationship may exist between production functions and cost functions.

Before returning to dual theory, let us digress for a moment to consider the function concept. This concept is one of the most important ideas of mathematics. Intuitively, we can characterize a function (**f**) as a **rule** which, given certain objects (**arguments**), will determine an object corresponding to them (**value**). Essential to a function is the uniqueness of the relationship it expresses between the **argument** (x) and **value** (y). Every argument of a function must be given one and only one value by the function. Different types of functions may be defined, for instance, "squaring" and "doubling" are functions taking positive integers into positive integers; the result of squaring a positive integer (n) is (n^2), and the result of doubling it is (2n). Another example is the relationship between factor inputs used by a farmer and the amount produced of its output is also functional, since to produce one unit of output the farmer needs to use some combination of the factor inputs. A farmer's production normally requires more than one argument (say land, seeds, fertilizer, etc.). This is an example of n-arguments functions.

Traditionally, people were content to think of functions as a pictorial, geometrical concept. This concept of functionality appeared to be too narrow. For instance, it does not allow one to handle discontinuities. Thus, mathematicians have tended to move away from the pictorial, geometrical concept of functions to a more **set-theoretical** characterization. Under this new approach a function may be derived by simply identifying it with its corresponding graph.[1] Generalizing these ideas, and making them explicit, we may define a function as follows:

> Let X and Z be sets. A function from X to Z (or a function with domain X and range Z) is a subset of the Cartesian Product XZ, such that for all $x \in X$, there is a unique member $z \in Z$ such that the ordered pair $<x,z> \in f$.

You may have noticed it is inessential that the range Z of a function **f** from X to Z be **exhausted**.[2] In the case where the range is exhausted, mathematicians often speak of **f** as being a function from x **onto** z. On the other hand, when the range is not exhausted

---

1 A graph of a function is a set of points in the Cartesian plane, which, in turn, is defined as the set R^2 of ordered pairs of real numbers.

2 The range of a function is exhausted when there may be members of $z \in Z$ such that for all $x \in X$, $f(x)=z$. In other words, every value of a function F is the image Z of only one argument X.

they speak of **f** as being a function from X **into** Z.

Notice that every value of a function need not be the image of only one argument. For instance take the function of squaring on the positive and non-positive integers. Every integer has a unique square, but not every square is the result of squaring only one integer. The number 4, for example, is the result of squaring either 2 or -2. This function does not have the property of being reversible. Many interesting functions, however, do have this property of "reversibility"; such functions are said to be **one-to-one** (or to be **bijections**).

The one-to-one function concept is relevant in duality theory. Duality, used in this context, means that all the information needed to derive a production function is contained in its corresponding optimal cost function, or viceversa. The necessary condition for a dual relationship between production and cost functions is that both be one-to-one functions. If both functions are monotonically increasing in their arguments then the necessary condition for these functions is immediately held. Notice, that the familiar neoclassical three-stage production function is not one-to-one, because two values of factors' vector correspond to at least some value of output. Only the first-stage of the neoclassical production function is relevant in duality theory because at this stage it is monotonically increasing in its arguments.

In a single factor setting, the duality of the production and the corresponding cost function is relatively simple. In such a case, we can characterize a production function (f) as a **rule** which relates inputs of factors (x) to the maximal output levels (y). That is

$$y=f(x) \tag{1}$$

If (f) is a one-to-one type (that is to say, every value of the function f(x) is the image of only one argument), and its inverse exists, then the corresponding cost function expressed in **physical** terms is the inverse of the production function (1):

$$x=f^{-1}(y), \tag{2}$$

where $\mathbf{f}^{-1}$ is the inverse of $\mathbf{f}$.

A simple example is the production function $y=x^b$. The corresponding dual cost function expressed in physical terms is $x=y^{1/b}$. All the information with respect to the parameter (b) of the production function is obtained from the corresponding dual cost function. Cost functions are usually expressed in dollar, rather than in physical terms. Under constant input price (p_x) assumption the cost function, expressed in dollar terms, is:

$$p_x x=p_x f^{-1}(y) \tag{3}$$

An important condition derived from duality theory is that, if any point on a single production function represents a technical maximum output (y) for a specific level of input use (x) associated with that point, then each point on the inverse cost function is optimal. That is, this point represents the lowest cost method of producing specific amount of output associated with the chosen point.

Notice, that if the underlying production function is not always monotonically increasing, the dual cost is not a one-to-one function. Thus, a point on the dual cost function is not necessarily a least cost point for the chosen level of output.

In a multifactor setting, the duality of the production function and its corresponding cost function becomes more complicated. McFadden (1978) specifies the production function's set conditions under which the corresponding dual cost function can exist. These conditions are:

1. **non-negative marginal products of the inputs.** The non-negativity implies free disposal of inputs. This assumption implies that if there is some input vector, denoted as x', which can produce some output vector called y', then a second bundle called x", which is at least as large as x', the x" can also produce y'. One implication of this assumption is that the isoquant maps consisting of concentric rings are ruled out, and that positively sloped isoquants are not allowed.
2. **non-increasing marginal rates of substitution between input pairs.** That is, in the two factor case, $\frac{d(dx_2/dx_1)}{dx^1}$ is non-positive. This implies that each isoquant is weakly convex to the origin.

If conditions (1) and (2) are met, then the minimum cost function corresponding to the production function will have the following properties.[3] It:

i. exists;

ii. is continuous;

iii. is non-decreasing for each price in the input price vector;

iv. is homogeneous of degree one in all variable input prices; implying that if all input prices double, so will all total variable costs; and

v. is concave in each input price for a given level of output (y^*).

---

[3] Detailed proof of these properties can be found in McFadden (1978, pp. 10-13).

A key characteristic of a particular class of production functions, known as **homothetic** production function, is that a line of constant slope (isocline) drawn from the origin of the corresponding isoquant map will connect points of the same slope. Here, the ratio of the factors remain fixed or constant, independent of the level of the output. The isocline represents the least cost combination of inputs, at given factor prices, at which the production level may be expanded, i.e., **the expansion path factor beam**. The production surface arising above this expansion path represents the minimum cost of producing a given level of output. The production function represented by the expansion path conditions (y^*) along the isocline in an n-input setting can be written as:

$$y^*=f(x^*), \tag{4}$$

where $x^*=[x^*_1, \ldots, x^*_n]$ the least cost quantities of $x_1, \ldots x_n$. The cost function that is dual to equation 4 (hereafter called **the indirect cost function**) can be obtained by making use of the expansion path conditions, and can be written as:

$$C^*=g(p,y^*) \tag{5}$$

where C^* is the least cost method of producing the output level y as defined by the expansion path conditions, given inputs price vector p.[4] The Marginal cost associated with the least cost marginal cost is:

$$MC^*=\frac{dC^*}{dy*}, \tag{6}$$

while the average cost associated with the least cost factor beam is

---

[4] For an example of this method see Appendix A.

$$AC^*=\frac{C^*}{y^*}. \tag{7}$$

The ratio of marginal to average cost along the least cost factor beam, or the dual cost elasticity Ψ^* that applies to the expansion path condition is:

$$\psi^*=\frac{1}{\epsilon}, \tag{8}$$

where $\in$ is the returns to scale parameter, or function coefficient for the underlying production function for the output arising from the least cost condition of inputs along the expansion path factor beam. If the total product along the expansion path is increasing at a decreasing, increasing, or constant rate, then costs are increasing at a decreasing, increasing, or constant rate, respectively.

Under competitive markets, the output price is a constant marginal revenue (MR). It will be equal to the least marginal cost (MC^*), only if MC^* is increasing with fixed input prices and primal production function is homogeneous.

The profit function representing the least cost method of generating a specific amount of profit (**the indirect profit function**), which corresponds to the dual cost function can be written as:

$$\pi^*=TR^*-C^*, \tag{9}$$

where TR represents total revenue of selling y^*, and C^* represents the indirect cost function of producing y^*.

Duality Theorems

The most famous theorems related to duality are **Hotelling's lemma** and **Shephard's lemma.** Both are specific applications of a mathematical theorem known as the envelope theorem. The envelope theorem, and the proofs of Hotelling's lemma, and Shephard's lemma are provided in Appendix B.

Shephard's lemma. Shephard's lemma states that a change in cost for the least (optimal) cost function with respect to the change in the price of the ith factor, evaluated at any particular level is equal to the ith factor that is used. More formally

$$\frac{\partial C^*}{\partial p_i}=x_i^*. \tag{10}$$

Hotelling's lemma. It states that a change in the indirect profit function arising from the **output expansion path** with respect to the kth product price (q_k) is equal to the optimal quantity of the kth output that is produced. That is

$$\frac{\partial \pi^*}{\partial q_k}=y_k^*. \tag{11}$$

Hotelling's lemma may also be applied to the factor side of production. It states that the change in the indirect profit function with respect to a change in the jth factor price is equal to the negative of the optimal quantity of the jth input as indicated by the **factor expansion path** condition. Thus we may write

$$\frac{\partial \pi^*}{\partial q_j} = -x_j^*. \tag{12}$$

Shephard's and Hotelling's lemmas are of considerable importance for empirical research. If the firm is operating according to the assumptions embodied in the expansion path conditions on both the factor and product sides, then product supply and factor demand equations can be obtained without any need of estimating the production function from physical input data.

II. The Cost Function, Scale and Scope Economies in Multiproduct Firms

The cost function is the single most useful tool in studying the economic behavior of a firm. Recent advances in duality theory and in empirical analysis techniques have guided economists away from the difficult task of directly estimating technological relationships (production functions and multi-output transformation functions) and toward the estimation of cost function.

First generation of cost function estimation studies virtually all dealt with **single-output** production, because the investigator either were attempting to provide useful simplification, or they did not recognize, or care to focus on the policies that arise only in the presence of **multi-output** production. Later, some investigators attempted to consider the presence of multi-output production by aggregating, in a fixed proportion, the different outputs of the firms into a single scalar measure over which costs of production may be "averaged". By doing so, in essence, these studies revert to the single-product case.

A simple example will provide more light on these arguments. Let us assume that the true cost relationship for an industry producing n different products, $y_i (i=1,...,n)$, is given by

$$C(y_1,...,y_n). \tag{13}$$

Let us define an aggregate "scalar output" as

$$y \equiv \sum_{i=1}^{n} \alpha_i y_i, \tag{14}$$

where α_i is the fixed output proportion of the aggregate scalar output. The estimation of the parameters of the associated scalar-output cost function implicitly requires the imposition of the following functional form:

$$\hat{C} = \hat{C}\left(\sum_{i=1}^{n} \alpha_i y_i\right). \tag{15}$$

The imposition of these functional form implicitly assumes strong separability[5] on its arguments. However, if this assumption does not hold then serious statistical biases will be introduced, rendering suspect any inference derived from equation (15).

Recent advances in industrial organization theory suggest that econometric examination of multi-product cost functions may yield insights into market structure and performance.[6] The replacement of a single measure of firm output by a set of disaggregated measures may substantially eliminate specification error. This should lead to more accurate estimates of the parameters of the aggregate scalar-output cost function. In addition, multi-product cost function estimates may be useful for answering additional questions concerning economies of scale, cost complementarities, economies of scope, natural monopoly, and optimal product mix.

The aim of this section is two fold. First, it attempts to discuss some properties of cost functions related with the single-product (or aggregate scalar product) industry.

---

[5] For details of separability concept see section below.

[6] See Baumol, Panzar & Willig (1982) for an extensive discussion of these advances.

Secondly, it embarks upon the quest for multi-product cost concepts that yield insight into multi-product properties both analogous and with no counterpart to those cost concepts of single-product case.

1. Single (or Aggregate Scalar) Product Cost Function Characteristics

Before starting the formal analysis of different cost concepts, it is importante to define one of the most basic cost concept:

Subadditivity

> Let y^i be the amount of the single-output produced by ith firm. A cost function is subadditive at total output $y(=\Sigma y^i)$ if it is more expensive for two or more firms to produce y than it is for a single firm to do so.

In other words, the cost of producing the whole is less than the sum of the costs of producing the parts. More formally,

$$C(y)<\sum_{i}^{k} C(y^i). \tag{16}$$

Subadditivity may be considered a local concept in the sense that costs may be subadditive to one output level, but not necessarily at another. However, when we want to determine whether costs are subadditive at a particular output level (y^*), we require a **global** concept of subadditivity. In this case, it is necessary to know the behavior of costs at other levels of operation below than those currently observed. That is, to know whether single-firm production of y^* is (or is not) cheaper to produce than **any** combination of smaller firms, one must know the magnitudes of the costs that **would** be incurred by **any** of the smaller firms. More formally, it is necessary ascertain

$$C(y^*)<\sum_{i=1}^{k} C(y^i),\ \textit{for every}\ y^*\leq y. \tag{17}$$

An important derivation from subadditivity concept, in the single-output case, is the natural monopoly concept.

Natural Monopoly

An industry is said to be a natural monopoly, in the single-output case, if over the entire relevant range of output the firm's cost is subadditive.

Average Cost (AC)

In the single-output case, average cost is defined by

$$AC(y)=\frac{C(y)}{y}. \tag{18}$$

Marginal Cost (MC)

In this case marginal cost is formally defined as

$$MC(y)=\frac{\partial C(y)}{\partial y}. \tag{19}$$

Baumol, Panzar & Willig (1982) have demonstrated the following propositions:

1. declining marginal costs through y imply declining average costs, however, the converse is not true. Note in Figure (1) that marginal cost is rising between v and

w, despite a declining average cost,

2. declining average cost implies subadditivity, but the converse is not true.[7]

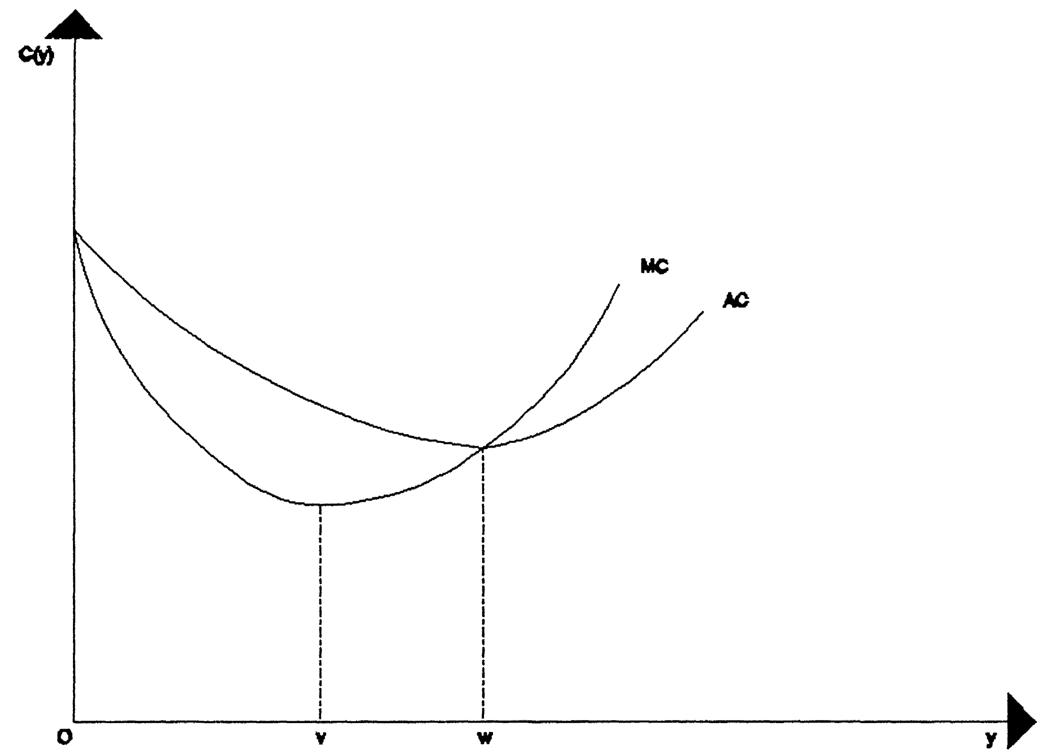

Figure 1 Average and Marginal Cost Curves in Single-Output Case.

(a) Traditional Returns to Scale Concept. Traditionally, returns to scale has been easily defined for homogeneous production function (Henderson & Quandt, 1971). Returns to scale are present, in this context, when an α-fold **proportional** increase in every input quantity (x) yields an α^k increase in output. For instance, in the single-output two-input case where k is a constant, and alpha is any positive real number. Returns to scale are increasing

[7] For a proof of this argument see BPW (1982); pp. 19-20.

$$f(\alpha x_1, \alpha x_2) = \alpha^k f(x_1, x_2), \tag{20}$$

if $k > 1$, constant if $k = 1$, and decreasing if $k < 1$. This definition of increasing returns to scale or economies of scale (i.e., $k > 1$) is impling that average cost will decline through output y. The reason is straightforward. If one wishes to increase any output y^* by the factor α^k ($k > 1$), the cheapest way to do so need not to be a proportionate increase in all inputs. Thus, even if average cost does not fall when output is increased by expanding all inputs proportionately, it may nevertheless fall when output is expanded in the most efficient manner, changing input proportions if appropriate. Under this definition economies of scale through output y imply that average cost will decline through output y, but not viceversa. If one wishes to increase a given level of output (y^*) by the factor α, the cheapest way to do so need not to be a proportionate increase in all inputs, it may fall when output is expanded by changing input proportions if appropriate, as well.

The ambiguities established above has been avoided by using a most recent definition of scale economies provided by BPW:

(b) Non-Traditional Returns to Scale Concept. The degree of returns to scale at y is given by the following relationship:

$$S = \frac{C(y)}{yMC(y)} = \frac{AC(y)}{MC(y)} \tag{21}$$

Returns to scale are increasing, constant, or decreasing as S is greater than, equal to, or less than unity. S corresponds to the output elasticity of output at y with respect to the cost incurred to produce it; $S = d\ln y/d\ln C(y)$. S is also the elasticity of output with respect to the cost of a proportionate expansion in all inputs, from any combination of input levels that is efficient for the production of y.

2. Multi-Product Cost Function Characteristics.

Some multi-output cost functions characteristics, such as ray average costs (RAC) and returns to scale, proceed in the same way for a single-output cost function. Moreover, other multi-output cost characteristics such as average incremental costs, product specific returns to scale, and economies of scope are multiproduct cost characteristics with no counterpart in the single-output cost case.

Ray average cost and returns to scales definition are based in the existence of a composite commodity. A composite good may be constructed by considering in an arbitrary way, the proportion in which the different products will be combined. Next, one must decide on the quantity of the bundle that will represent the value unity of the products, and thus one may measure the absolute quantity of the composite good.

Ray average cost (RAC)

Ray average cost of producing t units of the composite good y is defined to be

$$RAC\ (ty^{o}) = \frac{C\ (ty^{o})}{t}. \tag{22}$$

Geometrically, the ray average cost is measured by the slope of the line from the origin to

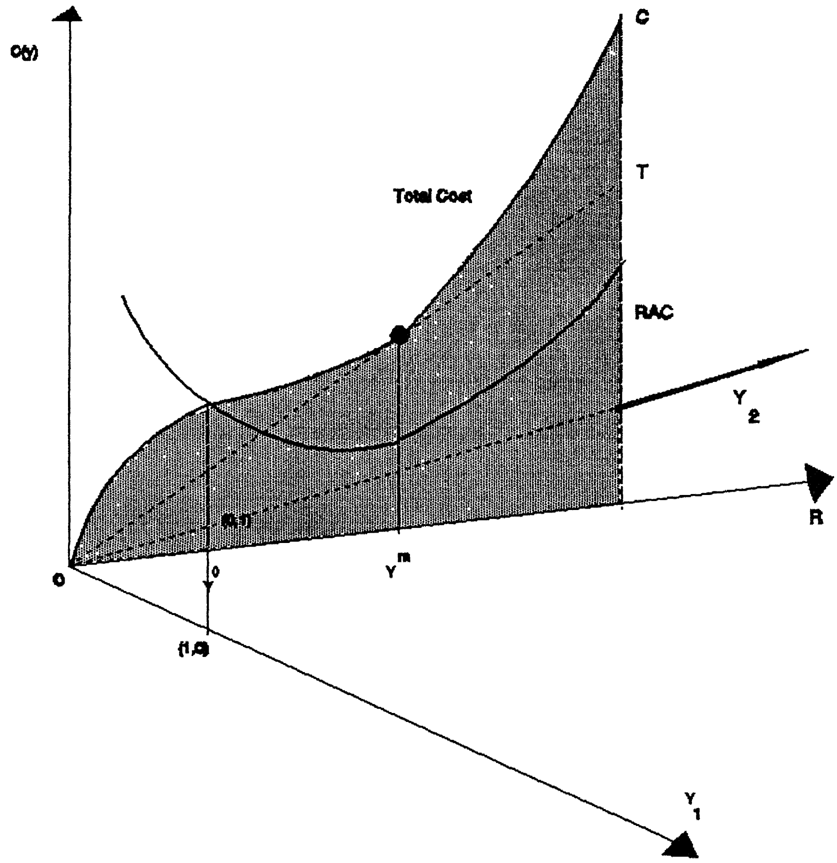

Figure 2 Ray Average Cost Curve.

any point on the cost surface (OC) along the ray (OR) (see Figure 2). In Figure 2 we see that the RAC and the total cost curves along ray OR have the usual relationships. They intersect at the unit level y^0 and RAC reaches its minimum at the output $y=y_m$, at which the ray OT is tangent to the total cost surface in the hyperplane erected on OR.

The RAC and multi-product returns to scale concepts relate to the proportional changes in the quantities in the entire product set. However, the magnitude of a firm operation may also change through variations in the output of one product holding the quantities of the other products constant. The cost of such variation may be considered as the incremental cost of ith product.

Incremental Cost

The incremental cost of the product i at y is

$$IC_i = C(y) - C(y_{N-i}), \tag{23}$$

where y_{N-1} is a vector with zero component in place y_i and components equal to those of y for the remaining products.

Having defined incremental cost we can then define average incremental costs.

Average incremental cost (AIC).

The average of incremental cost of product i is

$$AIC_i(y) = \frac{IC_i(y)}{y_i}. \tag{24}$$

Geometrically, this cost datum is found from the cross section of cost surface along portion of output surface (such as ST in Figure 3) that parallels to the axis for product y_2. In Figure 3 $T=(y^*_1, y^*_2)$ is a given output vector, S is the corresponding output vector on the y_1 axis at which y_1 has been held the same as at the point T, but y_2 has been reduced to zero. If product 2 has no output-specific fixed costs, then the total cost surface rises continuously above ST (curve AE). The height CE in Figure 3 measures the total incremental cost of product 2 at output vector T. The average incremental cost of product z, $AIC_2(y^*_1, y^*_2)$, is clearly given by the slope of the line from S to any point on the cost surface AE along the output surface ST.

It is clear that the average incremental cost of product 2 in Figure 3 are declining through y2, at least between O and y^*_2. This suggests, by analogy to the single-output case, the concept of product-specific scale economies.

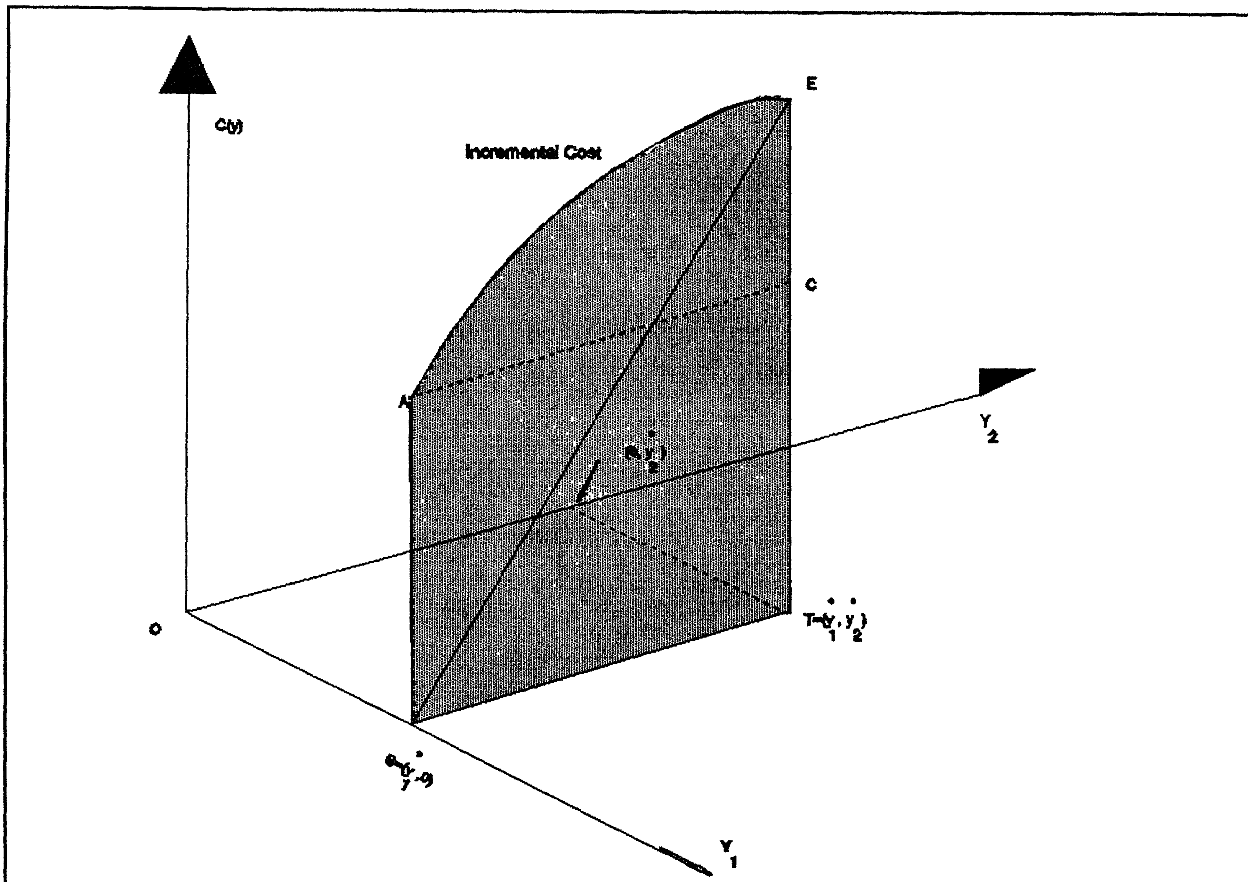

Figure 3 Average Incremental Cost.

Product-specific returns to scale. The degree of scale economies specific to product i at output vector y is given by

$$S_i(y) = \frac{IC_i(y)}{y_i C_i} \equiv \frac{AIC_i}{\partial C / \partial y_i}. \tag{25}$$

Returns to the scale of the product i at y are said to be increasing, decreasing, or constant if $S_i(y)$ is greater than, less than, or equal to unity, respectively.

In order to complete our definition of RAC, we still must select the unit of output

(y^o) for the composite good. Baumol, Panzar, and Willig find it useful to define the unit of output along each ray in terms of the distance from the origin to the unit simplex along that ray. Here the unit of simplex is given by

$$y^o=\left\{y\geq 0 \mid \sum_{i=1}^{n} y_i=1\right\}. \tag{26}$$

It is represented in Figure 2 by the 45° line with endpoints (1,0) and (0,1) on the axis. Then, the RAC(y) of producing the output vector $y \neq 0$ may be defined as

$$RAC(y)=\frac{C(y)}{\sum_{i=1}^{n} y_i}. \tag{27}$$

Ray average cost is said to be increasing (decreasing) at y if RAC(ty) is an increasing (decreasing) function of the scalar t, at t=1. RAC is said to be minimized at y if

$$RAC(y)<RAC(ty), \textit{ for all positive } t\neq 1. \tag{28}$$

Multi-product overall returns to scale

The degree of scale economies defined over the entire product set, N={1,...,n}, at y, is given by

$$S_N=\frac{C(y)}{y\nabla C(y)}\equiv\frac{C(y)}{\sum_i yC_i(y)}. \tag{29}$$

where $C_i(y) \equiv \partial C(y)/\partial y$. Returns to scale are said to be increasing, constant, or decreasing as S_N is greater than, equal to, or less than unity, respectively.

The degree of economies of scale and the elasticity of RAC(ty) with respect to t, (e), at the output point y are related by the following relationship

$$S_N = \frac{1}{1+e}. \tag{30}$$

Thus, returns to scale at the point y are increasing, decreasing, or locally constant ($S_N > 1, S_N < 1, S_N = 1$, respectively) as (e) is negative, positive, or zero, respectively.

Increasing multi-product returns to scale concept is strongest than declining RAC at y, because, it implies that RAC is decreasing at y. However, the converse is not true for the same reason as in the scalar case. For example, decreasing RAC at y does not imply that there must be increasing returns at y: RAC(ty) may be strictly declining at t=1 without having a negative derivative there. We can easily produce a natural multiproduct extension of our measure of product-specific returns to scale for the scalar case. The degree of scale economies specific to the product set T, subset of N, at y is given by

$$S_T(y) \equiv \frac{C_t(y)}{\sum_{j \in T} y_j C_j(y)}. \tag{31}$$

where $IC_T = C(y) - C(Y_{N-T})$, with y_{N-T} being a vector with zero components associated with the products in T and components equal in value to those of y for products in N-T (note that $y = y_T + y_{N-T}$). We can also state

$$S_T = \frac{1}{1+e_T}. \tag{32}$$

where e_T is the elasticity of average incremental cost of y for products in t.

Next, we shall describe the behavior of the cost surface along portions of output space that cut diagonally from one axis to another. In this case, we are considering the possibility that cost savings may result from simultaneous production of several different outputs in a single enterprise, as contrasted with their production in isolation, each by its own specialized firm. That is, there may exist economies resulting from the scope of the firm's operation.

Economies of scope

Let $P=\{T_1,...,T_k\}$ denote a non-trivial partition of S subset of N. That is, $U_iT_i=S$, $T_i \cap T_j=0$ for $i \neq j$, $T_i \neq 0$, and $k>1$. There are economies of scope at y_s with respect to the partition P if

$$\sum_{i=1}^{k} C(y_{T_i})>C(y_s). \tag{33}$$

If the inequality is reverted we may talk about **diseconomies of scope**.

Geometrically, the concept involves a comparison of $C(y_1^*,0)+C(0,y_2^*)$ in Figure 4, the sum of the heights of the cost surface over the corresponding points on the axis, with $C(y_1^*,y_2^*)$, the height of the cost surface at the point (y_1^*,y_2^*), which is the vector sum of $(0,y_1^*)$ and $(0,y_2^*)$. In Figure 4 the height of D above (y_1^*,y_2^*) must equal $C(y_1^*,0)+C(0,y_2^*)$.[8] The degree of economies of scope at y relative to the product set T may be defined more formally as

---

[8] Since the hyperplane may be described by

$C=ay_1+by_2$

where a and b are parameters. Therefore,

$C(y_1^*,0)=ay_1$, and $C(y_2^*,0)=by_2$.

Then, $C(y_1^*,y_2^*)$ must be less than $ay_1^*+by_2^*$ for economies of scope to hold.

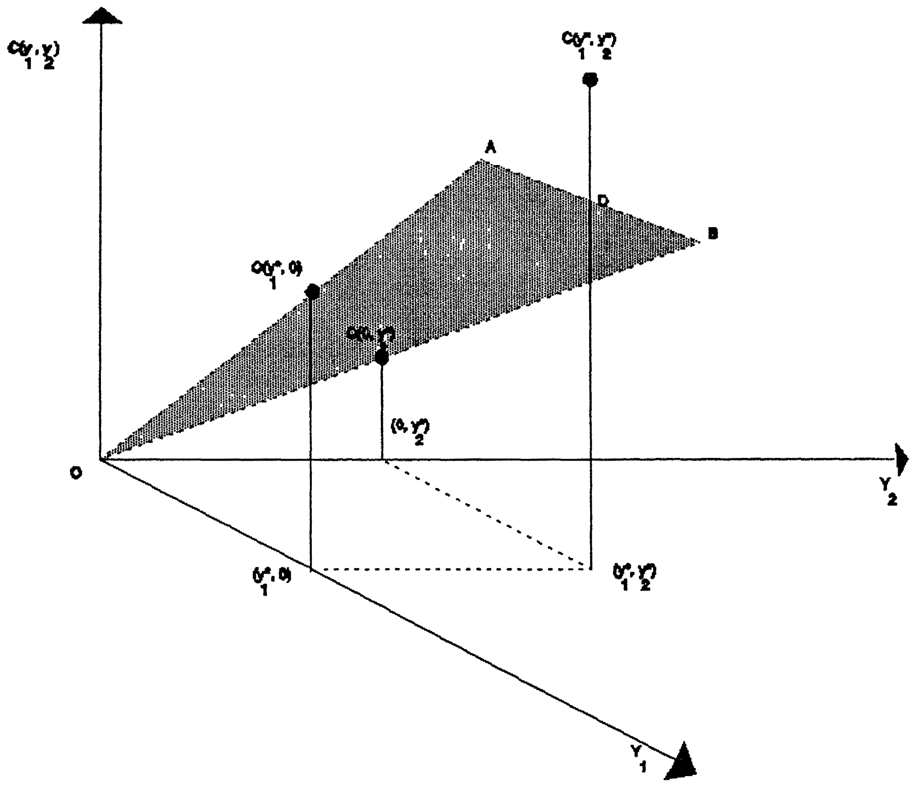

Figure 4 Economies of Scope.

$$SC_T(y) \equiv \frac{[C(y_T) + C(y_{N-T}) - C(y)]}{C(y)}. \tag{34}$$

III. The Elasticity of Substitution

The elasticity of substitution is a pure number that indicates the extent to which one input substitutes for another and hence indicates the shape of an isoquant according to the "usual" definition (Henderson and Quandt). The elasticity of substitution can be represented by the ratio of two percentages. Many expressions for the elasticity of substitution between two pair of factors has been widely discussed in the economic literature. In what follows we shall study the elasticity of substitution between pair of factors in two different setting. The two-factor setting and the n-factor setting.

Elasticity of Substitution in the Two-Factor Case

Suppose that there are two inputs, x_1 and x_2. The elasticity of substitution between x_1 and x_2 is usually defines as

$$\sigma = \frac{\textit{percent change in } \left(\frac{x_2}{x_1}\right)}{\textit{percent change in } MRS_{x_1 x_2}} \tag{35}$$

where $MRSx_1x_2$ represents the marginal rate of substitution between x_1 and x_2.

By equation (35), right angled isoquants (the classical example is tractor and the tractor's driver) have zero elasticity of substitution, while diagonal isoquants have an elasticity of substitution approaching infinity.

From equation (35) two approximately equivalent expressions for the elasticity of substitution between two input pairs of factors may be derived. These are the Arc Elasticity of Substitution and the Point Elasticity of Substitution.

The Arc Elasticity of Substitution. The Arc Elasticity of Substitution may be derived by substituting the percentage of change at period t respect period t-1 for the expression 'percent of change' (Δ) in equation (31). Thus equation (31) may be rewritten as

$$\sigma_A = \frac{\left[\frac{\Delta(x_2/x_1)}{(x_2/x_1)}\right]}{\left[\frac{\Delta MRS_{x_1x_2}}{MRS_{x_1x_2}}\right]}. \tag{36}$$

The arc elasticity of substitution represents the proportional percentage change in the input ratio (x_2/x_1) relative to the percentage change in the marginal rate of substitution.

As one move along an isoquant from point P_1 to P_2 in Figure 5, two things may happen. First, the ratio of the inputs (x_1/x_2) changes, Secondly, the slope of the isoquant, as measured by $MRSx_1x_2$, at point P_2 is different from its slope at point P_1. The ratio of these two changes, in percentages terms, is the arc elasticity of substitution.

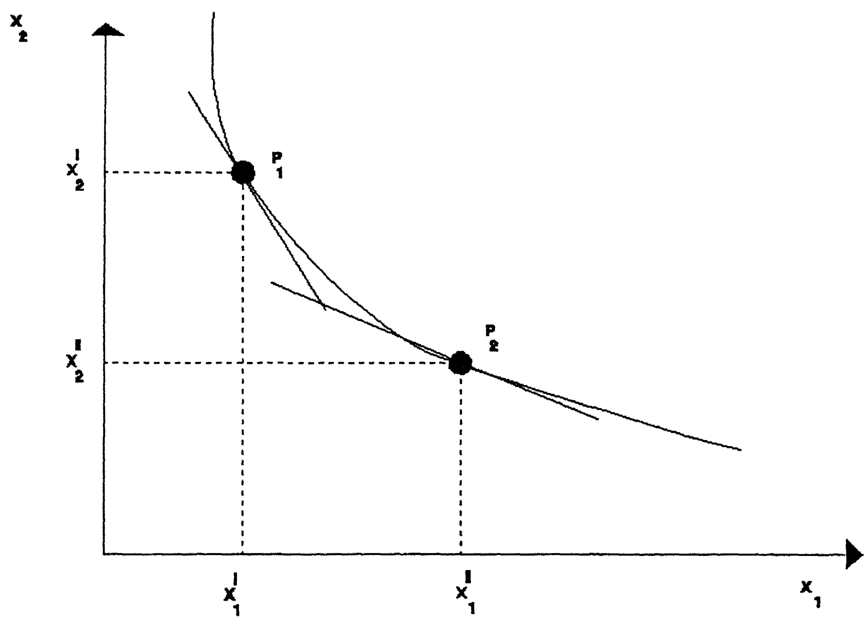

Figure 5 Graphical Representation of the Elasticity of Substitution.

Point elasticity of substitution. The point elasticity of substitution is defined by

$$\sigma_p = \frac{\left[\dfrac{d(x_2/x_1)}{(x_2/x_1)}\right]}{\left[\dfrac{dMRS_{x_1x_2}}{MRS_{x_1x_2}}\right]} = \frac{dln(x_2/x_1)}{dln(MRS_{x_1x_2})}, \qquad \textbf{(37)}$$

where the expression (ln) stands for natural logarithm.

If competitive markets are assumed, then the marginal rate of substitution between the pairs of factors equals the relative factors price (P_1/P_2), at the point of least cost combination on the isoquant. Thus, the point elasticity of substitution can be rewritten as:

$$\sigma_p = \frac{dln(x_1/x_2)}{dln(p_1/p_2)}. \tag{38}$$

Equation (38) is the elasticity of substitution attributed to Hicks (see also Varian 1984).

In the two-factor case, the elasticity of substitution will lie between zero and plus infinity. However, if more than two inputs are utilized, some input pairs may **complement** each other, leading to a potential negative elasticity for some of the input pairs.

Elasticity of substitution in the n-factor case.

The definition of the elasticity of substitution in an n-factor setting is further more complicated. In this case, a series of specific assumptions must be made with regard to the prices and input levels for those factors of production not directly involved in the elasticity of substitution calculation. As a result, the elasticity of substitution between inputs i and j will vary depending on these assumptions. In the n-factor case, a number of alternative definitions for the elasticity of substitution are also possible.

The one-input one-price elasticity of substitution. This type of elasticity of substitution may be defined as fixed proportion (β) of the cross price input demand elasticity evaluated at constant output:

$$\Phi_{ij} = \beta\left(\frac{dlnx_j}{dlnp_i}\right) \tag{39}$$

One example of elasticity of substitution of this kind is the *Hicks-Allen elasticity of substitution (HAES)*. Its form is a one-facto one-price elasticity of substitution since only one factor price (i) and one factor quantity (j) are involved:

$$\sigma_{ij}^{H}=S_j^{-1}\epsilon_{ij} \tag{40}$$

where s_j is the share of total cost attributed to the jth factor, p_jx_j/C^*, and ϵ_{ij} is the cross price factor demand elasticity evaluated at constant output, $dlnx_i/dlnp_j$. Notice also that the Hicks-Allen own price elasticity of substitution can be defined as:

$$\sigma_{jj}^{H}=s_j^{-1}\epsilon_{jj}, \tag{41}$$

where $s_j=p_jx_j/C^*$,and $\epsilon_{jj}=dlnx_j/dlnp_j$.

The two-input one-price elasticity of substitution. This type of elasticity of substitution involves two factor quantities but only one factor price:

$$\omega_{ij}=\frac{dln(x_j/x_i)}{dlnp_i}. \tag{42}$$

Each of these alternative definitions may be evaluated assuming constant output, cost, or marginal cost. Furthermore, it must be assumed that the prices on the remaining inputs other than i and j are held constant, or allowed to vary as p_i and p_j vary, which generates short and long run elasticity of substitution measures.

An extension of HAES is the Miroshima Elasticity of Substitution (Koizumi, 1976). It is an example of two-factor one-price elasticity of substitution. This elasticity of substitution is defined in terms of HAES as:

$$\sigma_{ij}^{M}=s_j\left(\sigma_{ij}^{H}-\sigma_{jj}^{H}\right)=\epsilon_{ij}-\epsilon_{jj}=\frac{dln(x_i/x_j)}{dlnp_j}. \tag{43}$$

Notice that the Miroshima elasticity of substitution is not symmetric, that is

$$\frac{dln(x_i/x_j)}{dlnp_j}\neq\frac{dln(x_j/x_i)}{dlnp_i}, \tag{44}$$

and therefore $\sigma_{ij}^{M}\neq\sigma_{ji}^{M}$.

The two-input two-price elasticity of substitution. An example of this type elasticity of substitution is the termed Shadow Elasticity of Substitution (McFadden, 1963). It allows all factors involved in the calculation to vary. Consequently, it can be considered as a long run elasticity of substitution. The shadow elasticity of substitution can be expressed in terms of HAES measure as

$$\sigma_{ij}^{S}=\left[\frac{s_i s_j}{s_i+s_j}\right]\left[2\sigma_{ij}^{H}-\sigma_{ii}^{H}-\sigma_{jj}^{H}\right]. \tag{45}$$

Thus, if the HAES and input cost share data ae available, the shadow elasticity of substitution can be readily calculated.

IV. FUNCTIONAL FORMS

Specific production functions used by researchers in empirical analysis frequently embody assumptions that are related with the functional forms itself. These assumptions have been referred in the economic literature as **maintained hypothesis.** These maintained hypotheses are not frequently recognized by the researcher, but do impose constraints on the possible outcome that can be generated by the analysis.

An example of a maintained hypothesis is provided by the **Cobb-Douglas production function (CD).** This production function specifies an aggregate neoclassical production function relating output to the aggregate inputs of capital and labor services.

$$y_t = AL_t^{\beta_1} K_t^{\beta_2}, \tag{46}$$

the underlying maintained hypothesis in CD type of specification is the hypothesis that the two-factor elasticity of substitution between any input pairs is constant and equal to 1. This holds even if the production is not linearly homogeneous, and $\beta_1 + \beta_2 \neq 1$. A simple proof is

$$MRS_{X_1X_2} = \beta X, \tag{47}$$

where $\beta = (\beta_1/\beta_2)$ and $X = (x_1/x_2)$.

$$lnMRS_{X_1X_2} = lnX + \ln\beta, \tag{48}$$

Thus,

$$lnX = lnMRS_{x_1x_2} - \ln\beta, \tag{49}$$

In an effort to avoid the maintained hypothesis regarding the elasticity of substitution

$$\sigma=\frac{dlnX}{dlnMRS_{x_1x_2}}=1. \tag{50}$$

of functional form of CD type Arrow et al (1961) specified the constant elasticity of substitution production function (CES) without the linear homogeneity property imposed. The CES is

$$y=A\left[\beta_1 x_1^{-\rho}+\beta_2 x_2^{-\rho}\right]^{-\frac{1}{\rho}}. \tag{51}$$

The elasticity of substitution is given by the power to which the factors are raised, $(1+\rho)^{-1}$. A simple proof is the marginal rate of substitution for the CES of the form

$$MRS_{x_1x_2}=\beta X^{(1+\rho)}, \tag{52}$$

where $\beta=(\beta_1/\beta_2)$ and $X=x_1/x_2$. Taking logarithms

$$lnMRS_{x_1x_2}=\ln\beta+(1+\rho)lnX, \tag{53}$$

$$lnX=\frac{lnMRS_{x_1x_2}-\ln\beta}{(1+\rho)}, \tag{54}$$

thus,

$$\sigma=\frac{dlnx}{dlnMRS_{x_1x_2}}=\frac{1}{(1+\rho)}. \tag{55}$$

The CES production function represents an appropriate improvement if the interest is centered on the elasticity of substitution within a production process using only two inputs, such as capital and labor. Nevertheless, extending the function to the n-input case, the maintained hypothesis that the same elasticity of substitution apply to every factor pair will still hold. Agricultural economic studies are usually interested in disaggregating input categories into more than two inputs. Thus, a more flexible functional form was needed for agricultural economic research.

An interesting and useful approach, referring to functional forms specification, has been provided by Diewert (1971). Diewert has recognized the close linkages that exists between various functional forms. He states that one way of looking at various functional forms is in terms of Taylor's series expansion. For example the Cobb-Douglas type of production function could be written as a first order Taylor's series expansion of lny in lnx_i:

$$lny=\alpha_o+\sum_{i=1}^{n}\beta_i lnx_i. \tag{56}$$

The CES, in turn, is a first order Taylor's expansion of y^ρ in x_i^ρ. In an n-factor setting, the CES could be written as

$$y^\rho=\alpha_o+\sum_{i=1}^{n}\beta_i x_i^\rho. \tag{57}$$

The transcendental logarithmic production function (translog for short), proposed by Christensen et al (1971, 1973), is simply a second order Taylor's expansion of lny in lnx_i:

$$lny=\alpha_o+\sum_{i=1}^{n}\beta_i lnx_i+1/2\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}lnx_i lnx_j. \tag{58}$$

The translog functional form has been widely utilized in agricultural economic research, because of the following important characteristics:

1. it is closely linked to the CD functional form. In fact, the translog is the CD when all β_{ij} are equal zero;
2. it is linear in the parameters, which makes the parameter estimation simple;
3. it is also normally monotonically increasing with respect to the use of each input under the usual parameter function;
4. it has no maintained hypothesis concerning the elasticity of substitution between any factor pairs;
5. it allows one to apply the fundamental concepts of duality.

V. FUNCTIONAL SEPARABILITY

Conventional production functions, such as multifactor CD and CES, are based in the specification of aggregate neoclassical production functions relating output to the aggregate inputs of capital and labor service (e.g., equipment and structures) or of different labor services (such as man-hours for production and non-production activities) into capital aggregates service or labor aggregates service, respectively. The use of such capital and labor aggregates, however, assumes that their different components could be **separated** into subfunctions. If such separability is possible, efficiency in production or consumption can be realized by **sequential optimization**. That is to say, production decisions, relative to factor intensities, can be optimized within each separable subset. Then the optimal use of factors can be obtained by optimizing the between-subset factors. Let us clarify these arguments with an example. Assume that the production function may be written as

$$y=F(K,L,N), \tag{59}$$

where K,L, and N are input aggregates of capital services, labor, and intermediate materials (e.g.,energy). Assume that it is permissible to separate the production function (59) in two subsets representing a value added index (V) of labor and capital services, and intermediate materials, N. Thus, equation (59) may be rewritten as

$$Y=F(V(K,L),N). \tag{60}$$

In such a context, the value added index (v) refers to the quantity of output produced per unit of time using capital and labor services. The optimal use of factors K, L, and N may be attained in a two-stage independent process. First, optimal labor/capital ratio (L/K) is obtained by optimizing the value added index, V(K,L). In the second stage of optimizing decisions the optimal N/V ratio is obtained by optimizing the gross production decisions, F(V,N).

The concept of separability has been established in a theoretical framework which consider a twice differentiable, strictly quasi-concave homothetic production function, with a finite number of inputs, each having a strictly positive marginal product.

$$y=F(x_1,x_2,\dots,x_n). \tag{61}$$

Assume that the set on n inputs is denoted N={1,...,n} and is **partitioned** into r mutually exclusive and exhaustive subsets N_s (s=1,2,...,r). This partition will be called R. The first and second partial derivatives of F(X) are denoted by F_i and F_{ij}.

$$F_i=\frac{\partial F}{\partial x_i}, \textit{ all input levels other than } x_i \textit{ held constant,} \tag{62}$$

with i=1,...,n. And

$$F_{ij}=\frac{\partial F_i}{\partial x_j}, \textit{ all input levels other than } x_i \textit{ and } x_j \textit{ held constant,} \tag{63}$$

with i,j=1,...,n

Weak Functional Separability

The production is said to be weakly separable with respect to the partition R if

where $MRSx_ix_j$ represents the marginal rate of substitution between inputs pair i and j. In other words, a production function is weakly separable with respect to R if the $MRSx_ix_j$ from

$$\frac{\partial MRS_{x_i x_j}}{\partial x_k}=0,\ \textit{for all } i,j \in N_s,\ \textit{and } k \notin N_s, \tag{64}$$

any subset N_s is independent of the quantities of inputs outside N_s.

Strong Functional Separability

The production function (59) is said to be strongly separable with respect to a partition R, if the MRS between any two inputs from subset N_s and N_k not depend on the quantities of inputs outside of N_s and N_t i.e.,

$$\frac{\partial MRS_{x_{ij}}}{\partial x_k}=0,\ \textit{for all } i \in N_s, j \in N_k, k \notin N_s \cup N_k. \tag{65}$$

Strong separability implies weak separability. However, weak separability implies strong separability only when the partition R is limited to two subsets.

Alternatively, the condition for factor i and j to be functionally separable from factor k is that the first and second derivatives of F satisfy

$$F_i F_{jk} - F_j F_{ik} = 0. \tag{66}$$

For weak separability this condition must hold for inputs i and j in one subset and input k in another subset. For strong separability this condition must hold in addition for inputs i, j, and k all in distinct subsets.

Berndt and Christensen (1973) established that separability restrictions on production and cost functions are equivalent to certain equality restrictions on the HAES. In a three-factor linear homogeneous production function, the following are equivalent restrictions on equation (59) at any point in input space:

1. factors L and K are functionally weakly separable from N,[9] i.e.,

$$F(K,L,N)=F(V(K,L),N). \tag{67}$$

2. equality of the HAES, i.e., $\sigma_{LN}=\sigma_{KN}$;
3. it exists a consistent aggregate price index, P^* and a consistent aggregate quantity index X^*, with components p_1 and p_2, and x_1 and x_2, respectively;
4. it exists a path independence of a Divisia Price index P^* and a Divisia Quantity index X^*;

From equation (67) three different types of separability may exist:

1. the separability of L and K from N. That is, the first and second derivatives must satisfy

$$F_L F_{KN}-F_K F_{LN}=0 \tag{68}$$

or

$$\sigma_{LN}=\sigma_{KN}. \tag{69}$$

2. the separability of L and N from K. The first and second derivatives must satisfy

---

[9] Notice that, since the partition is limited to two subsets, weak and strong separability are equivalent restrictions.

$$F_L F_{NK} - F_N F_{LK} = 0 \tag{70}$$

or

$$\sigma_{LK} = \sigma_{NK} \tag{71}$$

3. the separability conditions, in this case, are satisfied if and only if

$$F_K F_{NL} - F_N F_{KL} = 0 \tag{72}$$

or

$$\sigma_{KL} = \sigma_{NL}. \tag{73}$$

It is clear that only two of these three types of separability conditions are independents. For instance, $\sigma_{LN} = \sigma_{KN}$ and $\sigma_{LK} = \sigma_{NK}$ implies that $\sigma_{KL} = \sigma_{NL}$.

VI. THE FACTOR PRICE FRONTIER

This section provides a detailed exposition of the concept of factor price frontier (FPF) in a two and three-factor economy.

The Factor Price Frontier in the Two-Factor Case

Any description of the economic performance of an individual firm, an industry, a sector of the economy, or the economy as a whole must start with a production function relating output or the product to the input of factor of production. In the two-factor case, the production function relates the conventional pair of inputs, labor (L) and capital (K), with the value-added output (V) at a given level of technology (T). We can write the output as

$$V=G(L,K;T). \tag{74}$$

Assume that the production function (74) is linear homogeneous. Thus, the following properties hold

1. $MP_L=V_L=\partial G/\partial L$, $MP_K=V_K=\partial G/\partial K$, where MP_L and MP_K denote the marginal productivities of labor and capital, respectively,
2. increasing the inputs by a certain factor of proportionality (α) output increases by the same factor, that is (excluding T),

$$\alpha V = G(\alpha L, \alpha K). \tag{75}$$

Under assumption of constant returns to scale only the relative proportions of labor, capital, and output matter for the marginal products. Therefore, the dimensionality of production is reduced from 3 to 2. We may, for instance, divide output and capital by labor (say $\alpha = 1/L$) and rewrite equation (74) as

$$\frac{V}{L} = G\left(1, \frac{K}{L}\right) = g\left(\frac{K}{L}\right). \tag{76}$$

Alternatively, if we consider ($\alpha = 1/V$), we may rewrite equation (70) as

$$1 = G\left(\frac{L}{V}, \frac{K}{V}\right). \tag{77}$$

This gives the unit isoquant G in Figure 5.

The curve G describes all the efficient combinations of labor (L) and capital (K) that yield one unit of output. The slope of curve G at any given choice of factor use [(L/V)o,(K/V)o], at the production point P_o, measures the ratio of marginal products (marginal rate of substitution between L and K, at the given level of factor use), V_L/V_K.

Under Euler's law, linear homogeneous production function satisfy the following condition:

$$V_L L + V_K K = V(K, L), \tag{78}$$

or

$$V_L\frac{L}{V}+V_K\frac{K}{V}=1. \tag{79}$$

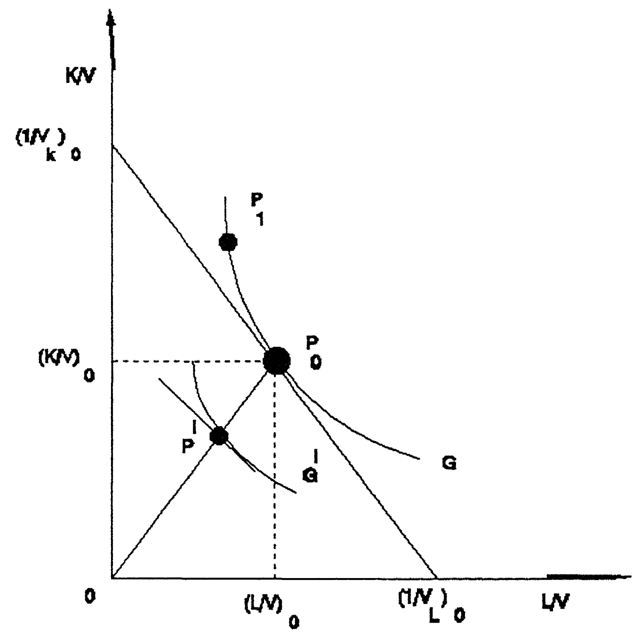

Figure 6 Production (the Primal).

Thus, the intercepts of the tangent on the two axes in figure 5 measure the respective reciprocals of the marginal products. These intercepts, under cost minimization, also represent the marginal cost of production expressed in the units of each factor.[10] Another convenient property of the linear homogeneous production function is given by the fact that the relative distances of the pairs of points along the axes measure the respective product elasticities (or shares) of labor and capital,

$$s=\frac{L/V}{1/V_L}=\frac{V_L L}{V}, \quad 1-s=\frac{K/V}{1/V_K}. \tag{80}$$

We can also see from the diagram that the point P_o divides the tangent (and likewise the points $(L/V)_o$, and $(K/V)_o$ divide the respective intercept) by the ratio of the factor shares s/(1-s). This ratio must sum up to one, a well known-property of linearly homogeneous production function (Euler's theorem).

When producers not only minimize production costs for a given level of output, but also choose their output competitively so as to maximize profits, marginal costs equal the product price level, P_v. Thus, factor shares may be rewritten as

$$s=\frac{WL}{P_V V}, \; 1-s=\frac{RK}{P_V V}, \tag{81}$$

In sum, under constant returns to scale, the curve G in Figure 5 summarizes all there is to know about the production technology. It establishes a one-to-one correspondence between the pair of marginal products and the pair of unit inputs and, thus, also with the

---

[10] Under cost minimizing behavior, producers will choose the input combination which satisfies

$$W/V_L=Z/V_K=MC,$$

where W and Z denote the respective marginal factor cost of producing one unit of V.

capital/labor ratio. This ratio, in turn, is represented by the slope of the ray OP_o to the production point, P_o. Under diminishing returns (the curve G is convex to the origin), a higher capital/labor ratio at point P_1 goes together with higher K/V ratio, lower L/V ratio, and likewise with lower MP_k and higher MP_L.

This simple correspondence between intensity of input use and marginal factor products, termed "duality", allows one to describe the productive process interchangeably in the space of marginal factor products. The resulting curve F, in Figure 3, is the factor price frontier (FPF).

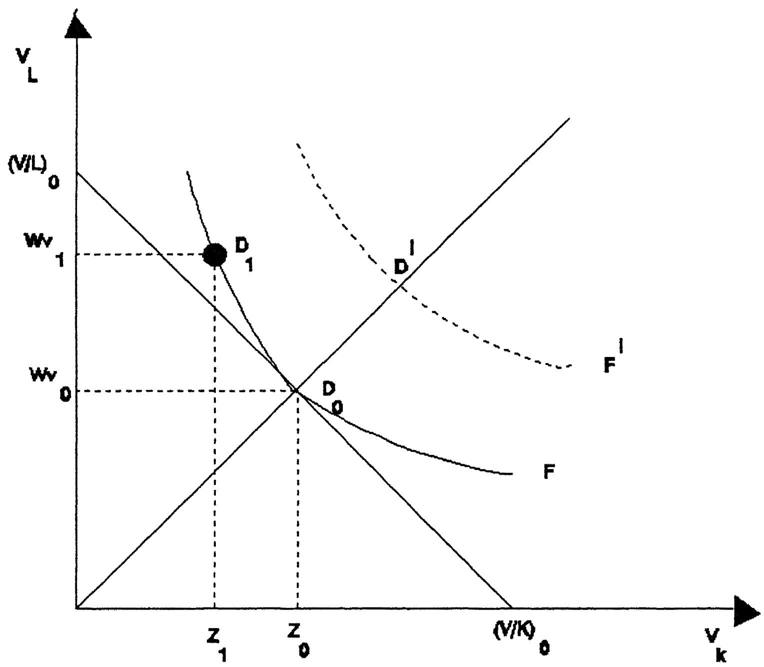

Figure 7 Factor Price Fontier (the Dual).

The factor price frontier. The FPF is the dual of the isoquant G (in Figure 5) and represents the pair of maximal combination of marginal factor products. If profits are maximized and factors are paid their real marginal products ($MP_L = W/P_V$ and $MP_K = Z/P_V$) then the FPF represents the maximum rate of return that can be paid to capital for a given real wage level, independent of the level of activity.

REFERENCES

Arrow, K., H.B. Chenery, B. Menhald, and R.M.Solow, "Capital Labor Substitution and Economic Efficiency," Review of Economics and Statistics, 43, 1961.

Baumol, W., J. Panzar and R. Willig, Contestable Markets and the Theory of Industry Structure, New York: Harcourt Brace Jovanovich, 1982.

Beattie, B.R., and C.R. Taylor, The Economics of Production, New York: Wiley, 1985.

Berndt, E.R. and L. Christensen, "The Translog Function and the Substitution of Equipament, Structures, and Labor in U.S. Manufacturing 1929-68," Journal of Econometrics, 1, 1973.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau, "Conjute Duality and Transcendental Logarithmic Production Function," Econometrica, 39, 1971.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau, "Transcendental Logarithmic Production Frontiers," Review of Economic and Statistics, 5, 1973.

Diewert, W.E., "An application of the Shephard Duality Theorem; A Generalized Leontief Production Function," Journal of Political Economy, 79, 1971.

Henderson, J.M., and R.E. Quandt, Microeconoic Theory: A Mathematical Approach, 2nd. Ed. New York: McGraw Hill, 1971.

Koizumi, T., "A Further Note on the Definition of Elasticity of Substitution in the Many Input Case," Metroeconomica, 28, 1976.

McFadden, Daniel, "Constant Elasticity of Substitution Production Functions," Review of Economic Studies, 30, 1963.

McFadden, Daniel, "Cost, Revenue and Profit Functions," in M.Fuss and D.McFadden eds., Production Economics: A Dual Approach to Theory and Application, Vol. 1, Amsterdam: North Holland, 1978.

Varian, H., Microeconomic Analysis, New York: W.W. Norton, 2d edition, 1984.

APPENDIX A

ESTIMATION OF AN INDIRECT COST FUNCTION

Suppose that the production function is given by

$$y=x_1^{\alpha_1}x_2^{\alpha_2}, \tag{A.1}$$

where α_1 and α_2 are parameters, x_1 and x_2 are inputs, and p_1 and p_2 are the respective input prices.

The input cost function is:

$$C=p_1x_1+p_2x_2. \tag{A.2}$$

The estimation of the indirect cost function of production function (A.1) is defined as follows:

Step 1. Find the respective expansion path. This is done by partially differentiating the production function (A.1) with respect to x_1 and x_2, to find the respective marginal products. Then, the negative ratio of the marginal products (the marginal rate of technical substitution, $MRSx_1x_2$) is equated to the inverse input price ratio (p_1/p_2). Thus, the relationship that defines the least combination along the expansion path may be expressed as:

$$\alpha_1p_2x_2=\alpha_2p_1x_1. \tag{A.3}$$

Step 2. Determine the quantity of inputs, as defined by the expansion path conditions, that are used in terms of cost (A.2) and parameters of the production function (A.1), and input prices. Solve (A.3) for x_1:

$$x_1=\left[\frac{\alpha_1 p_2 x_2}{\alpha_2 p_1}\right]. \tag{A.4}$$

Substituting (A.4) into (A.2) and factoring x_2 out:

$$x_2=C\left(\alpha_1 p_2 \alpha_2^{-1}+p_2\right)^{-1} \tag{A.5}$$

Similarly, for X_1:

$$x_1=C\left(\alpha_2 p_1 \alpha_1^{-1}+p_1\right)^{-1}. \tag{A.6}$$

Step 3. Define the production function in terms of the expansion path conditions. Substitute (A.5) and (A.6) into the original production function (A.1) and rearranging terms:

$$y=C^{\alpha_1+\alpha_2}\left(\alpha_2 p_1 \alpha_1^{-1}+p_1\right)^{-\alpha_1}\left(\alpha_1 p_2 \alpha_2^{-1}+p_2\right)^{-\alpha_2} \tag{A.7}$$

Step 4. Define the total cost function that is dual to the production function defined along the expansion path factor beam (indirect cost function).

Solving (A.7) for C in terms of y, the production function parameters and the input prices yields the optimal cost function of producing the specific output level y as defined by the expansion path conditions:

$$C=y^{(1/\alpha_1+\alpha_2)}\left(\alpha^{-1}\alpha_2 p_1+p_1\right)^{(\alpha_1/\alpha_1+\alpha_2)}\left(\alpha_2^{-1}\alpha_1 p_2+p_1\right)^{(\alpha_2/\alpha_1+\alpha_2)}$$

$$=y^{1/(\alpha_1+\alpha_2)} p_1^{\alpha_1/(\alpha_1+\alpha_2)} p_2^{\alpha_2/(\alpha_1+\alpha_2)}$$

$$=y^{1/(\alpha_1+\alpha_2)} Z \qquad \text{(A.8)}$$

Notice that the value of Z is treated as constant, since it is dependent only on the assumed constant prices of the inputs and the assumed constant parameters of the production function. Thus, any point on the dual cost function (A.8) representing a particular quantity of output (y) is optimal in the sense that it represents the least cost combinations of inputs needed to produce y.

Notice also that at most only one point on the dual cost function represents global optimality, where the marginal cost (MC^*) of producing the incremental unit of output using the least cost combination of factors is exactly equal to the marginal revenue (MR^*) obtained from producing the incremental unit of $\hat{y}$.

APPENDIX B

The Envelope Theory

Following Beattie and Taylor (1985), let

$$Z=g(w_1,\ldots,w_n;\alpha) \tag{B.1}$$

be a function to be maximized with respect to each w_i for a given parameter vector α. The first order conditions

$$\frac{\partial g}{\partial w_i}=0 \tag{B.2}$$

for $i=1,\ldots,n$,

define the optimal value (w_i^*) for each w_i in terms of the parameter vector α. That is

$$w_i^*=w_i^*(\alpha) \tag{B.3}$$

Thus, the optimal value for (B.1) is:

$$Z^*=g(w_1^*,\ldots,w_n^*;\alpha) \tag{B.4}$$

The envelope theorem states that the rate of change in Z^* with respect to a change in α, if all w_i are allowed to adjust, is equal to the change in g with respect to the change in the parameter α when all w_i are assumed to be constant.

More formally, the envelope theorem may be stated as:

$$\frac{\partial Z^*}{\partial \alpha}=\frac{\partial g}{\partial \alpha} \tag{B.5}$$

Proof:

Partially differentiating (B.4) with respect to α gives

$$\frac{\partial Z^*}{\partial \alpha}=\sum_{i=1}^{n}\left(\frac{\partial g}{\partial w_i^*}\right)\left(\frac{\partial w_i^*}{\partial \alpha}\right)+\frac{\partial g}{\partial \alpha} \tag{B.6}$$

Now, if condition (B.2) holds, then the first term of the right-hand side of (B.6) must be equal to zero and equation (B.5) holds.

The Shephard's Lemma

Consider the case of a cost minimizing firm using n different inputs (x) in order to produce a given level of output ($\hat{y}$). The Lagrangian for minimizing costs subject to the constraint imposed by the production function is:

$$L=\sum_{i=1}^{n} p_i x_i+\lambda[\hat{y}-f(x_1,...,n)] \tag{B.7}$$

The corresponding first order conditions are:

$$\frac{\partial L}{\partial x_i}=p_i-\lambda\frac{\partial f}{\partial x_i}=0 \tag{B.8}$$

That is, the optimal x_i (x_i^*) is obtained from

$$p_i=\lambda\frac{\partial f}{\partial x_i}, \tag{B.9}$$

for $i=1,...,n$.

Thus, the indirect cost function (C^*), representing the least cost way of producing y, given price of inputs is:

where x_i^* represent the quantities of inputs defined by the expansion path factor beam.

$$C^{*}=\sum_{i=1}^{n} p_i x_i^{*} \tag{B.10}$$

The Shephard's lemma states that the rate of change in the indirect cost function with respect to the change in the price of the ith factor, evaluated at any particular output level, is equal to the optimal quantity of the ith indicated by the factor expansion path condition.

That is

$$\frac{\partial C^{*}}{\partial p_i}=x_i^{*}, \tag{B.11}$$

for i = 1,...,n.

Proof:

Partially differentiating (B.10) with respect to the ith factor price yields:

$$\frac{\partial C^{*}}{\partial p_i}=\sum_{i=1}^{n} p_i \frac{\partial x_i^{*}}{\partial p_i}+x_i^{*} \tag{B.12}$$

Substituting (B.9) into (B.12):

$$\frac{\partial C^{*}}{\partial P_i}=\sum_{i=1}^{n} \lambda \frac{\partial f_i}{\partial C_i} \frac{\partial x_i^{*}}{\partial p_i}+x_i^{*} \tag{B.13}$$

Now assume that the original production function is defined at the cost minimizing level of input use:

$$y=f(x_1^{*},\ldots,x_n^{*}) \tag{B.14}$$

Maximizing the production function with respect to a change in the ith input price:

for i = 1,...n.

$$\frac{\partial y}{\partial p_i}=\frac{\partial f}{\partial x_i^*}\cdot\frac{\partial x_i^*}{\partial p_i}=0 \tag{B.15}$$

Substituting (B.15) into (B.13) and Shephard's lemma (B.11) holds.

The Hotelling's Lemma

Consider the case of a firm using n different inputs (x) in order to produce m different outputs (y). Define total revenue (R) as

$$R=\sum_{j=1}^{m} q_j y_j \tag{B.15}$$

where q_j is the price of the jth output. Total cost is defined as

$$C=\sum_{i=1}^{n} p_i x_i \tag{B.17}$$

The indirect revenue function (R^*) represents the optimal allocation of outputs to maximize revenue, and can be specified as

$$R^*=\sum_{j=1}^{m} q_j y_j^* \tag{B.18}$$

The corresponding indirect cost function (C^*) is

$$C^*=\sum_{i=1}^{n} p_i x_i^* \tag{B.19}$$

Thus, the indirect profit function may be defined as

The Hotelling's lemma shows that a change in the profit function arising from the

$$\pi^*=R^*-C^*=\sum_{j=1}^{m} q_j y_j-\sum_{i=1}^{n} p_i x_i^* \tag{B.20}$$

output expansion path (the indirect profit function) with respect to the kth product price or input price, is equal to the kth output that is produced,

$$\frac{\partial\pi^*}{\partial q_k}=y_k^*, \tag{B.21}$$

for k = 1,...,m,

or to the negative of the kth input that is used

$$\frac{\partial\pi^*}{\partial p_i}=-x_k^*, \tag{B.22}$$

for k = 1,...,n.

Implicitly, the profit maximizing transformation production function may be written as

$$F(y_1^*,\dots,y_m^*;x_1^*,\dots,x_n^*)=0 \tag{B.23}$$

The profit maximizing Lagrangian is:

$$L=\sum_{j=1}^{m} q_j y_j-\sum_{i=1}^{n} p_i x_i+\eta\left[f(y_1,\dots,y_m;x_1,\dots,x_n)-0\right] \tag{B.24}$$

The corresponding first order conditions, on the product side, are:

$$\frac{\partial L}{\partial y_j}=q_j+\eta\frac{\partial F}{\partial y_j}=0 \tag{B.25}$$

that is,

$$q_j = -\eta \frac{\partial F}{\partial y_j} \tag{B.26}$$

for $j=1,...,m$.

The optimal y_j is y_j^*.

The first order conditions on the factor side are

$$\frac{\partial L}{\partial x_i} = -p_i + \eta \frac{\partial F}{\partial x_i} = 0 \tag{B.27}$$

That is,

$$p_i = \eta \frac{\partial F}{\partial x_i} \tag{B.28}$$

for $i=1,...,n$.

The optimal x_i is x_i^*.

Partially differentiating (B.20) with respect to the kth product price:

$$\frac{\partial \pi^*}{\partial q_k} = y_k^* + \sum_{j \neq k}^{m} q_j \left(\frac{\partial y_j^*}{\partial q_k} \right) - \sum_{i=1}^{n} p_i \left(\frac{\partial x_i^*}{\partial q_k} \right) \tag{B.29}$$

Substituting (B.26) and (B.28) into (B.29):

$$\frac{\partial \pi^*}{\partial q_k} = y_k^* + \eta \left[\sum_{j \neq k}^{m} \left(\frac{\partial F}{\partial y_j^*} \right) \left(\frac{\partial y_j^*}{\partial q_k} \right) - \sum_{i=1}^{n} \left(\frac{\partial F}{\partial x_i^*} \right) \left(\frac{\partial x_i^*}{\partial q_k} \right) \right] \tag{B.30}$$

Partially differentiating (B.23) with respect to the kth product price:

$$\frac{\partial F}{\partial q_k}=\sum_{j=1}^{m}\left(\frac{\partial F}{\partial y_j^*}\right)\left(\frac{\partial y_j^*}{\partial q_k}\right)+\sum\left(\frac{\partial F}{\partial x_i^*}\right)\left(\frac{\partial x_i^*}{\partial q_k}\right)=0 \tag{B.31}$$

Substituting (B.31) into (B.30) and the Hotelling's lemma as applied to product side (B.21) holds.

Differentiate the indirect profit function with respect to the kth input price:

$$\frac{\partial \pi^*}{\partial p_k}=\sum_{j=1}^{m} q_j\left(\frac{\partial y_j^*}{\partial p_k}\right)-\sum_{i\neq k}^{n} p_i\left(\frac{\partial x_i^*}{\partial p_k}\right)-x_k^* \tag{B.32}$$

Substituting (B.20) and (B.28) into (B.32):

$$\frac{\partial \pi^*}{\partial p_k}=\eta\left[\sum_{j=1}^{m}\left(\frac{\partial F}{\partial y_j^*}\right)\left(\frac{\partial y_j^*}{\partial p_k}\right)-\sum_{i\neq k}^{n}\left(\frac{\partial F}{\partial x_i^*}\right)\left(\frac{\partial x_i^*}{\partial p_k}\right)\right]-x_k^* \tag{B.33}$$

differentiate (B.23) with respect to the kth input price:

$$\frac{\partial F}{\partial p_k}=\sum_{j=1}^{m}\left(\frac{\partial F}{\partial y_j^*}\right)\left(\frac{\partial y_j^*}{\partial p_k}\right)+\sum_{i=1}^{n}\left(\frac{\partial F}{\partial x_i^*}\right)\left(\frac{\partial x_i^*}{\partial p_k}\right)=0 \tag{B.34}$$

Substituting (B.34) into (B.30) and the Hotelling's lemma as applied to input side (B.22) holds.